

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2005, 06:56:21 ; Search time 42.1298 Seconds

(without alignments)
1349.183 Million cell updates/sec

Title: US-09-939-537-37

Perfect score: 591

Sequence: 1 TRFSSRAEPAYQQGQGNQLY.....LSTATQDYDALHMQALPPR 111

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: uniprot_03:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	572	96.8	163	1	CD3Z_PIG	Q9XJ99 sus scrofa
2	561.5	95.0	164	1	CD3Z_HUMAN	P20969 homo sapien
3	496	83.9	165	1	CD3Z_RABIT	Q9NUF8 corycolagus
4	476.5	80.6	164	1	CD3Z_MOUSE	P24161 mus musculus
5	471.5	79.8	166	1	CD3Z_SHEEP	P29329 ovie aries
6	378.5	64.0	206	1	CD3H_MOUSE	P29020 mus musculus
7	303	51.3	322	2	Q6KAV0	Q6KAV0 mus musculus
8	90.5	15.3	612	2	Q7SYB5	Q7SYB5 brachydanio
9	90.5	15.3	628	2	Q6TNU8	Q6TNU8 brachydanio
10	86	14.6	1167	1	WCI1_NEUCR	Q01371 neurospora
11	86	14.6	1262	2	Q7RYA7	Q7RYA7 neurospora
12	85	14.4	172	2	Q8CMD5	Q8CMD5 streptococc
13	81.5	13.8	744	2	Q754G7	Q754G7 ashbya gosw
14	81	13.7	1133	2	Q7XK73	Q7XK73 oryza sativ
15	80	13.5	692	2	Q19579	Q19579 caenorhabdi
16	79.5	13.5	868	2	Q9VYI5	Q9VYI5 drosophila
17	78.5	13.3	312	2	Q6SLR3	Q6SLR3 bacillus th
18	78.5	13.3	501	2	Q6U061	Q6U061 oryza sativ
19	78	13.2	919	2	Q9LPD8	Q9LPD8 arabidopsis
20	78	13.2	937	2	Q9MAL4	Q9MAL4 arabidopsis
21	77.5	13.1	328	2	Q6RGA4	Q6RGA4 rhodococcus
22	77	13.0	889	2	Q6H4V8	Q6H4V8 oryza sativ
23	77	13.0	928	2	Q6H4V9	Q6H4V9 oryza sativ
24	76.5	12.9	521	1	DB45_DROME	Q07886 drosophila
25	75.5	12.8	169	2	Q8RF77	Q8RF77 fusobacteri
26	75.5	12.8	169	2	Q8RG33	Q8RG33 fusobacteri
27	75.5	12.8	169	2	Q8RHV6	Q8RHV6 fusobacteri
28	75.5	12.8	924	2	Q92QJ4	Q92QJ4 rhizobium m
29	75	12.7	171	2	Q9KX22	Q9KX22 streptococc
30	75	12.7	354	2	Q91E80	Q91E80 cydia pomon
31	74.5	12.6	596	2	Q6PA69	Q6PA69 xenopus lae

32	74.5	12.6	761	2	Q7MNX6	Q7MNX6 vibrio vuln
33	74.5	12.6	761	2	Q8DEI4	Q8DEI4 vibrio caro
34	74	12.5	136	2	Q39681	Q39681 daucus caro
35	74	12.5	364	2	Q6C3D3	Q6C3D3 yarrowia li
36	74	12.5	450	2	Q8VB87	Q8VB87 mus musculus
37	74	12.5	748	2	Q95TP4	Q95TP4 drosophila
38	74	12.5	820	2	Q8CD02	Q8CD02 mus musculus
39	74	12.5	867	2	Q8IKI7	Q8IKI7 plasmodium
40	74	12.5	1330	2	Q8K4P0	Q8K4P0 mus musculus
41	74	12.5	3190	2	Q01368	Q01368 drosophila
42	74	12.5	3276	2	Q9WJ21	Q9WJ21 drosophila
43	73.5	12.4	122	1	GS1B_BACSU	P26907 bacillus su
44	73.5	12.4	360	1	RLPA_YERPE	Q84D96 yersinia pe
45	73.5	12.4	360	2	Q6BDF1	Q6BDF1 yersinia pe

ALIGNMENTS

RESULT 1	CD3Z_PIG	STANDARD;	PRT;	163 AA.
AC	Q9XSJ9;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	T-cell surface glycoprotein CD3 zeta chain precursor (T-cell receptor			
DE	T3 zeta chain).			
GN	Name=CD3Z;			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sub.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Minesota miniature swine;			
RA	Jie H.-B., Yim D., Kim Y.B.;			
RT	"The molecular cloning of porcine CD3 zeta.";			
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: Probable role in assembly and expression of the TCR			
CC	- SUBUNIT: The TCR/CD3 complex of T lymphocytes consists of either a			
CC	TCR alpha/beta or TCR gamma/delta heterodimer coexpressed at the			
CC	cell surface with the invariant subunits of CD3 labeled gamma,			
CC	delta, epsilon, zeta, and eta (By similarity).			
CC	- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	- PTM: Phosphorylated on Tyr residues after T-cell receptor			
CC	- SIMILARITY: Belongs to the CD3Z/FCER1G family.			
CC	- SIMILARITY: Contains 3 ITAM domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF153630; AAD34640.1; -			
DR	InterPro; IPR003110; ITAM.			
DR	Pfam; PF02189; ITAM; 3.			
DR	SMART; SM00077; ITAM; 3.			
KW	Phosphorylation; Receptor; Repeat; Signal; T-cell; Transmembrane.			
FT	CHAIN	1	21	
FT	STGNL	22	163	
FT	FT			
FT	DOMAIN	22	30	
FT	TRANSMEM	31	51	
FT	DOMAIN	52	163	
FT	DOMAIN	69	89	
FT	DOMAIN	107	128	
FT	DOMAIN	138	158	
FT	DOMAIN			
FT	ITAM 1.			
FT	Cytoplasmic (Potential).			
FT	ITAM 2.			
FT	ITAM 3.			

FT DISUPID 32 32 Interchain (Potential).
 FT MOD RES 152 152 Phosphotyrosine (By similarity).
 SQ SEQUENCE 163 AA; 18568 MW; 34898620B67167C7 CRC64;
 Query Match 96.8%; Score 572; DB 1; Length 163;
 Best Local Similarity 97.3%; Pred. No. 1.2e-46;
 Matches 107; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 RFSRSAPPPYOGGONOLYNELNGRREYDVLDKRGGRPEMGKPRKRNQEGLYNEL 61
 DB 54 KFRSRSDAPFYOGGONOLYNELNGRREYDVLDKRGGRPEMGKPRKRNQEGLYNEL 113
 QY 62 QKDKMAVAVSEIGMKRRRGKHDGLYQGLSTATQDTYDALHMOALPPR 111
 DB 114 QKDKMAVAVSEIGMKRRRGKHDGLYQGLSTATQDTYDALHMOALPPR 163
 RESULT 2
 ID CD32_HUMAN STANDARD; PRT; 164 AA.
 AC P20963; O8TX4;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE T-cell surface glycoprotein CD3 zeta chain precursor (T-cell receptor
 T3 zeta chain).
 GN Name=CD32; Synonyms=TCZ, TCRZ;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN 1
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89071765; PubMed=2974162;
 RA Weisman A.M., Hou D., Orloff D.G., Modi W.S., Senanez H.,
 RA O'Brien S.J., Klausner R.D.;
 RT "Molecular cloning and chromosomal localization of the human T-cell
 RT receptor zeta chain: distinction from the molecular CD3 complex";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9709-9713 (1988).
 RN 2
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Dietzenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo W.F., Cassavant T.L., Scheetz T.E.,
 RA Brownstein M.U., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mallaly S.J.,
 RA Bosak S.A., Mcwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Heiton E., Kettelman M., Madan A., Kodlygus S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield Y.S.N., Krzywinski M.I., Skalek U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN 3
 RP INTERACTION WITH HIV-1 NEF.
 RX PubMed=10224289;
 RA Xu X.-N., Lafert B., Sreaton G.R., Kraft M., Wolf D., Kolanus W.,
 RA Wongkolapap U., McMichael A.J., Baur A.S.;
 RT "Induction of Fas ligand expression by HIV involves the interaction of
 RT Nef with the T cell receptor zeta chain";
 RL J. Exp. Med. 189:1489-1496 (1999).
 RN 4
 RP INTERACTION WITH SLA.

RX MEDLINE=99380595; PubMed=10449770; DOI=10.1073/pnas.96.17.9775;
 RA Tang J., Sawaedhkosol S., Chang J.-H., Burakoff S.J.;
 RT "SLAP, a dimeric adapter protein, plays a functional role in T cell
 RT receptor signaling";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:9775-9780 (1999).
 RN 5
 RP INTERACTION WITH DOCK2.
 RX MEDLINE=22165501; PubMed=12176041; DOI=10.1016/S0006-291X(02)00931-2;
 RA Nishihara H., Maeda M., Tsuda M., Makino Y., Sawa H., Nagashima K.,
 RA Tanaka S.;
 RT "DOCK2 mediates T cell receptor-induced activation of Rac2 and IL-2
 RT transcription";
 RL Biochem. Biophys. Res. Commun. 296:716-720 (2002).
 RN 6
 RP PHOSPHORYLATION SITES TYR-83; TYR-111; TYR-123 AND TYR-142.
 RX PubMed=12522270; DOI=10.1073/pnas.2436191100;
 RA Salomon A.R., Ficarro S.B., Brill L.M., Brinker A., Phung Q.T.,
 RA Ericson C., Sauer K., Brock A., Horn D.M., Schultz P.G., Peters E.C.;
 RT "Profiling of tyrosine phosphorylation pathways in human cells using
 RT mass spectrometry";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:443-448 (2003).
 RN 7
 RP STRUCTURE BY NMR OF 136-149.
 RX MEDLINE=93201600; PubMed=7680960; DOI=10.1016/0092-8674(93)90405-F;
 RA Wakeman G., Shoelson S.E., Pant N., Cowburn D., Kuriyan J.;
 RT "Binding of a high affinity phosphotyrosyl peptide to the Src SH2
 RT domain: crystal structures of the complexed and peptide-free forms";
 RL Cell 72:779-790 (1993).
 CC -1- FUNCTION: Probable role in assembly and expression of the TCR
 CC complex as well as signal transduction upon antigen triggering.
 CC -1- SUBUNIT: The TCR/CD3 complex of T lymphocytes consists of either a
 CC TCR alpha/beta or TCR gamma/delta heterodimer coexpressed at the
 CC cell surface with the invariant subunits of CD3 labeled gamma,
 CC delta, epsilon, zeta, and eta. CD3-zeta forms either homodimers or
 CC heterodimers with CD3-eta. Interacts with SLA and SLA2. Interacts
 CC with DOCK2. Interacts with HIV-1 Nef protein.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=CD-3-zeta;
 CC IsoId=P20963-1; Sequence=Displayed;
 CC Name=CD-3-eta;
 CC IsoId=P20963-2; Sequence=Not described;
 CC -1- PTM: Phosphorylated on Tyr residues after T-cell receptor
 CC triggering (By similarity).
 CC -1- SIMILARITY: Belongs to the CD32/PCERIG family.
 CC -1- SIMILARITY: Contains 3 ITAM domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, J04132; AAA60394.1; -.
 CC EMBL, BC025703; AAH25703.1; -.
 CC PIR, A31768; A31768.
 CC PDB, ITCE; NMR; B=136-149.
 CC GeneW; HGNC:1677; CD32.
 CC H-InvDB; HIX001296; -.
 CC MIM; 186780; -.
 CC GO; GO:0005886; C:Plasma membrane; TAS.
 CC GO; GO:0042101; C:T-cell receptor complex; TAS.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC GO; GO:0042803; F:protein homodimerization activity; NAS.
 CC InterPro; IPR003110; ITAM.
 CC Pfam; PF02189; ITAM; 3.
 CC SMART; SM00077; ITAM; 3.
 CC 3D-structure; Alternative splicing; Phosphorylation; Receptor; Repeat;
 CC Signal; T-cell; Transmembrane.
 CC SIGNAL 1
 CC 21

```

FT CHAIN 22 164 T-cell surface glycoprotein CD3 zeta
FT DOMAIN 22 30 chain.
FT TRANSMEM 31 51 Extracellular (Potential).
FT DOMAIN 52 164 Potential.
FT DOMAIN 69 89 Cytoplasmic (Potential).
FT DOMAIN 108 129 ITAM 1.
FT DOMAIN 139 159 ITAM 2.
FT DISULFID 32 32 Interchain (Potential).
FT MOD_RES 83 83 Phosphotyrosine.
FT MOD_RES 111 111 Phosphotyrosine.
FT MOD_RES 123 123 Phosphotyrosine.
FT MOD_RES 142 142 Phosphotyrosine.
FT MOD_RES 153 153 Phosphotyrosine.
FT CONFLICT 60 61 DA -> EP (in Ref. 1).
FT CONFLICT 101 101 Missing (in Ref. 1).
SQ SEQUENCE 164 AA; 18696 MW; 9408260374856B9 CRC64;

Query Match 95.0%; Score 561.5; DB 1; Length 164;
Best Local Similarity 96.4%; Pred. No. 1.2e-45;
Matches 107; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 2 RRSRSAPPAAYQGQGNQLYNELNGRREBYDVKRGRDPENGGKRP-RRKNPQEGLYNE 60
DB 54 KFSRSADAPAYQGQGNQLYNELNGRREBYDVKRGRDPENGGKRPQRRKNPQEGLYNE 113
QY 61 LQDKMAEAYSEIGMKGERRGKHDGLYQGLSTATKDTYDALHMQALPPR 111
DB 114 LQDKMAEAYSEIGMKGERRGKHDGLYQGLSTATKDTYDALHMQALPPR 164

RESULT 3
CD3Z_RABIT STANDARD; PRT; 165 AA.
ID CD3Z_RABIT
AC Q9TUF8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE T-cell surface glycoprotein CD3 zeta chain precursor (T-cell receptor
T3 zeta chain).
GN Name=CD3z;
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B/J x Chbb.HM;
RA Isono T., Nishimura M.;
RT "Rabbit CD3 zeta."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probable role in assembly and expression of the TCR
CC complex as well as signal transduction upon antigen triggering.
CC -1- SUBUNIT: The TCR/CD3 complex of T lymphocytes consists of either a
CC TCR alpha/beta or TCR gamma/delta heterodimer coexpressed at the
CC cell surface with the invariant subunits of CD3 labeled gamma,
CC delta, epsilon, zeta, and eta (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: Phosphorylated on Tyr residues alter T-cell receptor
CC triggering (By similarity).
CC -1- SIMILARITY: Belongs to the CD3/FCER1G family.
CC -1- SIMILARITY: Contains 3 ITAM domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-eb.ch/announce/
CC or send an email to license@isb-eb.ch).
CC -----
CC EMBL; AB035152; BAA86994.1; -
CC InterPro; IPR003110; ITAM.

```

```

DR Pfam; PF02189; ITAM; 3.
DR SMART; SM00077; ITAM; 3.
KW Phosphorylation; Receptor; Repeat; Signal; T-cell; Transmembrane.
FT SIGNAL 1 21 By similarity.
FT CHAIN 22 165 T-cell surface glycoprotein CD3 zeta
FT DOMAIN 22 30 chain.
FT TRANSMEM 31 51 Extracellular (Potential).
FT DOMAIN 52 165 Potential.
FT DOMAIN 69 89 Cytoplasmic (Potential).
FT DOMAIN 107 128 ITAM 1.
FT DOMAIN 140 160 ITAM 2.
FT DISULFID 32 32 Interchain (Potential).
FT MOD_RES 154 154 Phosphotyrosine (By similarity).
SQ SEQUENCE 165 AA; 18773 MW; 3183136130BAAAF5 CRC64;

Query Match 83.9%; Score 496; DB 1; Length 165;
Best Local Similarity 84.8%; Pred. No. 2.2e-39;
Matches 95; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

QY 2 RRSRSAPPAAYQGQGNQLYNELNGRREBYDVKRGRDPENGGKRP-RRKNPQEGLYNE 61
DB 54 KFSRGEDEVSPQGHQTQLYNELNGRREBYDVKRGRDPENGGKRPQRRKNPQEGLYNE 113
QY 62 QDKMAEAYSEIGMKGE--RRRGKHDGLYQGLSTATKDTYDALHMQALPPR 111
DB 114 QDKMAEAYSEIGMKGENRRRGKHDGLYQGLSATKDTYDALHMQALPPR 165

RESULT 4
CD3Z_MOUSE STANDARD; PRT; 164 AA.
ID CD3Z_MOUSE
AC P24161; O903G3;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE T-cell surface glycoprotein CD3 zeta chain precursor (T-cell receptor
T3 zeta chain).
GN Name=CD3z; Synonyms=Ncrz;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC MEDLINE=88145643; PubMed=3278377;
RA Weissman A.M., Banlyash M., Hou D., Samelson L.E., Burgess W.H.,
RA Klausner R.D.;
RT "Molecular cloning of the zeta chain of the T cell antigen receptor."
RL Science 239:1018-1021 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89327299; PubMed=2787796;
RA Banlyash M., Heu V.W., Seldin M.F., Klausner R.D.;
RT "The isolation and characterization of the murine T cell antigen
RT receptor zeta chain gene."
RL J. Biol. Chem. 264:13252-13257 (1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=822354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Mikado I., Oseto N., Saito K., Suzuki H., Yamana H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schirm L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Bruste V., Choctha C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmerand T., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedziarski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,

```

RA Maglott D.R., Maltale L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Nunata K., Okido T., Pavan W.J., Petrea G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravezi T., Reed J.C., Reed D.J., Reid J., Ringwald M.,
 RA Savelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wymah-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carinci P., Hayatsu N.,
 RA Hironaka-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arikawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Yoshino M., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Hematopoietic;
 RX MEDLINE=2238825; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein W.J., Ustin T.B., Toshiyuki S., Carinci P., Prange S.J.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Malek J.A., Gunaratne P.H.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Walek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M.,
 RA Butterfield J.S.N., Krzywicki M.I., Skalska J., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=90239005; PubMed=2139725;
 RA Jin Y.J., Clayton L.K., Howard F.D., Koyasu S., Sieh M.,
 RA Steinbrich R., Tarr G.B., Reinherz E.L.,
 RT "Molecular cloning of the CD3 zeta subunit identifies a CD3 zeta-
 RT related product in thymus-derived cells."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3319-3323(1990).
 RN [6]
 RP ALTERNATIVE SPLICING.
 RX PubMed=2150596;
 RA Ohno H., Saito T.,
 RT "CD3 zeta and eta chains are produced by alternative splicing from a
 RT common gene."
 RL Int. Immunol. 2:1117-1119(1990).
 RN [7]
 RP ERRATUM.
 RA Ohno H., Saito T.,
 RN Int. Immunol. 4:1339-1339(1992).
 RN [8]
 RP INTERACTION WITH SLA.
 RX MEDLINE=20130290; PubMed=10662792;
 RA Sosinowski T., Pandey A., Dixit V.M., Weiss A.,
 RT "Src-like adaptor protein (SLAP) is a negative regulator of T cell
 RT receptor signaling."
 J. Exp. Med. 191:463-474(2000).
 RN [9]
 RP INTERACTION WITH SLA2.
 RX MEDLINE=22013997; PubMed=11891219; DOI=10.1074/jbc.M10318200;
 RA Pandey A., Ibarrola N., Kratchmarova I., Fernandez M.M.,
 RA Constantinescu S.N., Ohara O., Sawasdkosol S., Lodish H.F., Mann M.,

RT "A novel Src homology 2 domain-containing molecule, Src-like adaptor
 RT protein-2 (SLAP-2), which negatively regulates T cell receptor
 RT signaling."
 RL J. Biol. Chem. 277:19131-19138(2002).
 CC -1- FUNCTION: Probable role in assembly and expression of the TCR
 CC complex as well as signal transduction upon antigen triggering.
 CC -1- SUBUNIT: Interacts with DOK2 (By similarity). The TCR/CD3 complex
 CC of T lymphocytes consists of either a TCR alpha/beta or TCR
 CC gamma/delta heterodimer coexpressed at the cell surface with the
 CC invariant subunits of CD3 labeled gamma, delta, epsilon, zeta, and
 CC eta. CD3-zeta forms either homodimers or heterodimers with CD3-
 CC eta. Interacts with SLA and SLA2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=CD3-zeta;
 CC IsoId=P24161-1; Sequence=Displayed;
 CC Name=CD3-eta;
 CC IsoId=P29020-1; Sequence=External;
 CC -1- PTM: Phosphorylated on Tyr residues after T-cell receptor
 CC triggering.
 CC -1- SIMILARITY: Belongs to the CD3z/FCER1g family.
 CC -1- SIMILARITY: Contains 3 ITIM domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; M19729; AAA40171.1; -;
 DR EMBL; J04967; AAA50301.1; -;
 DR EMBL; AK017904; BAB30997.1; -;
 DR EMBL; BC052824; AAH52824.1; -;
 DR PIR; A40104; A40104.
 DR MGI; 88334; C03z.
 DR InterPro; IPR003110; ITIM.
 DR Pfam; PF02189; ITIM; 3.
 KW Alternative splicing; Direct protein sequencing; Phosphorylation;
 KW Receptor; Repeat; Signal; T-cell; Transmembrane.
 FT FT 1 21
 FT STGNAL
 FT CHAIN 22 164
 FT FT 22 30
 FT DOMAIN 31 51
 FT TRANSMEM 52 164
 FT FT 69 89
 FT DOMAIN 108 129
 FT FT 139 159
 FT DOMAIN 32 32
 FT DISULFID 153 153
 FT MOD RES 153 153
 FT CONFLICT 153 153
 FT SEQUENCE 164 AA; 18637 MW; 1B802035A312831 CRC64;
 SQ
 Query Match 80.6%; Score 476.5; DB 1; Length 164;
 Best Local Similarity 83.8%; Pred. No. 1.5e-37;
 Matches 93; Conservative 5; Mismatches 12; Indels 1; Gaps 1;
 Oy 2 RFSRSAPPAYOGGOLVYELNLGRREYDVLDRGRGDPENMGK-PRKQFQEGLYNE 60
 Db 54 KFSRSAPETANLDPPNLYELNLGRREYDVLDRGRGDPENMGKQQRRRNFBGVYNA 113
 Oy 61 LQDKMAEAYSEIGMKGRGGRGGLYGLSTARKQTYDNLHMAALPR 111
 Db 114 LQDKMAEAYSEIGTKGRGGRGGLYGLSTARKQTYDNLHMAALPR 164
 RESULT 5
 CD3Z_SHEEP STANDARD; PRT; 166 AA.
 ID CD3Z_SHEEP
 AC P29329;

DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE T-cell surface glycoprotein CD3 zeta chain precursor (T-cell receptor
 T3 zeta chain).
 GN Name=CD3z; Synonyms=CD3h;
 OS Ovis aries (Sheep).
 OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Caprinae; Ovis.
 NCBI_TaxId=9940;
 RX MEDLINE=9311305; PubMed=8420837;
 RA Hein W.R., Tunncliffe A.;
 RT "Invariant components of the sheep T-cell antigen receptor: cloning of
 the CD3 epsilon and Tcr zeta chains.";
 RL Immunogenetics 37:279-284(1993).
 CC -1- FUNCTION: Probable role in assembly and expression of the TCR
 complex as well as signal transduction upon antigen triggering.
 CC -1- SUBUNIT: The TCR/CD3 complex of T lymphocytes consists of either a
 TCR alpha/beta or TCR gamma/delta heterodimer coexpressed at the
 cell surface with the invariant subunits of CD3 labeled gamma,
 delta, epsilon, zeta, and eta.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PTM: Phosphorylated on Tyr residues after T-cell receptor
 triggering (By similarity).
 CC -1- SIMILARITY: Belongs to the CD3z/FCER1G family.
 CC -1- SIMILARITY: Contains 3 ITAM domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z12968; CA78312.1; -
 DR PIR: I46424; I46424.
 DR InterPro: IPR003110; ITAM.
 DR Pfam: PF02189; ITAM; 3.
 DR SMART: SM00077; ITAM; 3.
 KM Phosphorylation; Receptor; Repeat; Signal; T-cell; Transmembrane.
 FT SIGNAL 1 21 By similarity.
 FT CHAIN 22 166 T-cell surface glycoprotein CD3 zeta
 chain.
 FT DOMAIN 22 30 Extracellular (Potential).
 FT TRANSMEM 31 51 Potential.
 FT DOMAIN 52 166 Cytoplasmic (Potential).
 FT DOMAIN 69 89 ITAM 1.
 FT DOMAIN 108 129 ITAM 2.
 FT DOMAIN 141 161 ITAM 3.
 FT DISULFID 32 32 Interchain (Potential).
 FT MOD RES 155 155 Phosphotyrosine (By similarity).
 SQ SEQUENCE 166 AA; 18704 MW; E7D89AD84E58311A CRC64;

 Query Match 79.8%; Score 471.5; DB 1; Length 166;
 Best Local Similarity 80.5%; Pred. No. 4.7e-37;
 Matches 91; Conservative 10; Mismatches 9; Indels 3; Gaps 2;

 QY 2 RPSRSAPRAYOGGQNLVNLGRRREYDV.DKRGGRPEVGGKRP-RKKNQEGLYNE 60
 DB 54 KFRSRSDAPYQGNPNVNLVNGRREYAVLDKRGGRPEVGGKRPQRKKNHEVYNE 113
 QY 61 LQDKNAEAYSEIGMK--GERRRGKHGHDGLYOGISTATKDTYALHMOALPPR 111
 DB 114 LRDKNAEAYSEIGMSKSDNRRRGKHGHDGYOGISTATKDTYALHMOALPPR 166
 RESULT 6
 CD3H_MOUSE

ID CD3H_MOUSE STANDARD; PRT; 206 AA.
 AC P29020;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE T-cell surface glycoprotein CD3 eta chain precursor (T-cell receptor
 T3 eta chain).
 GN Name=CD3z; Synonyms=CD3h;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RX MEDLINE=90239005; PubMed=2139725;
 RA Jin Y.J., Clayton L.K., Howard F.D., Koyasu S., Sieh M.,
 RA Steindrich R., Tarr G.E., Reinherz E.L.;
 RT "Molecular cloning of the CD3 eta subunit identifies a CD3 zeta-
 related product in thymus-derived cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3319-3323(1990).
 RN [2]
 RP SEQUENCE OF 144-206 FROM N.A.
 RX MEDLINE=91271358; PubMed=1828894;
 RA Clayton L.K., D'Adamo L., Sieh M., Hussey R.E., Koyasu S.,
 RA Reinherz E.L., Howard F.B.;
 RT "CD3 eta and CD3 zeta are alternatively spliced products of a common
 genetic locus and are transcriptionally and/or post-transcriptionally
 regulated during T-cell development.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5202-5206(1991).
 RN [3]
 RP SEQUENCE OF 144-206 FROM N.A.
 RX PubMed=2150596;
 RA Ohno H., Saito T.;
 RT "CD3 zeta and eta chains are produced by alternative splicing from a
 common gene.";
 RL Int. Immunol. 2:1117-1119(1990).
 RN [4]
 RP ERRATUM.
 RA Ohno H., Saito T.;
 RL Int. Immunol. 4:1339-1339(1992).
 CC -1- FUNCTION: Probable role in assembly and expression of the TCR
 complex as well as signal transduction upon antigen triggering.
 CC -1- SUBUNIT: The TCR/CD3 complex of T lymphocytes consists of either a
 TCR alpha/beta or TCR gamma/delta heterodimer coexpressed at the
 cell surface with the invariant subunits of CD3 labeled gamma,
 delta, epsilon, zeta, and eta. CD3-eta can be complexed in a
 heterodimeric form with CD3-zeta subunit. CD3-eta homodimer has
 not been observed.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=CD-3-eta;
 CC IsoId=P29020-1; Sequence=Displayed;
 CC Name=CD-3-zeta;
 CC IsoId=P24161-1; Sequence=External;
 CC -1- SIMILARITY: Belongs to the CD3z/FCER1G family.
 CC -1- SIMILARITY: Contains 3 ITAM domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M33158; AAA37398.1; -
 DR EMBL: M76711; AAA40403.1; -
 DR PIR: A35900; A35900.
 DR MGI: 88334; CD3Z.
 DR InterPro: IPR003110; ITAM.
 DR Pfam: PF02189; ITAM; 2.
 DR SMART: SM00077; ITAM; 2.

```

KW Alternative splicing; Direct protein sequencing; Receptor; Repeat;
KM Signal; T-cell; Transmembrane.
FT CHAIN 1 21
FT 22 206 T-cell surface glycoprotein CD3 eta
FT DOMAIN 22 30 chain.
FT TRANSMEM 31 51 Extracellular (Potential).
FT DOMAIN 52 206 Potential.
FT DOMAIN 52 206 Cytoplasmic (Potential).
FT DOMAIN 69 89 ITAM 1.
FT DOMAIN 108 129 ITAM 2.
FT DOMAIN 139 159 ITAM 3.
FT DISULFID 32 32 Interchain (Potential).
SQ SEQUENCE 206 AA; 23339 MW; 829256A2CF4E444 CRC64;

Query Match 64.0%; Score 378.5; DB 1; Length 206;
Best Local Similarity 82.2%; Pred. No. 4.3e-28;
Matches 74; Conservative 5; Mismatches 10; Indels 1; Gaps 1;

QY 2 RFSRSAPPPAYQOGQNLVNEINLGRREYDVLDRKRGDPENMGK-PRKRPQEGLYNE 60
DB 54 KFSRSAPETANIQDPQNLVNEINLGRREYDVEKKRRADPEMGKQGRRRNPQEGVYNA 113
QY 61 LQDKMAEYSEIGMKGERRGKGDGLYQ 90
DB 114 LQDKMAEYSEIGTKGERRRGKGDGLYQ 143

RESULT 7
Q6KAVO PRELIMINARY; PRT; 322 AA.
AC Q6KAVO;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ46519.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RA Niinomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK128376; BC87407.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004888; F:transmembrane receptor activity; IEA.
DR GO; GO:0007166; P:cell surface receptor linked signal transdu. .; IEA.
DR InterPro; IPR003110; ITAM.
DR Pfam; PF02189; ITAM; 1.
DR SMART; SM00077; ITM; 1.
SQ SEQUENCE 322 AA; 34926 MW; D4461DBEC26BC011 CRC64;

Query Match 51.3%; Score 303; DB 2; Length 322;
Best Local Similarity 60.5%; Pred. No. 1.1e-20;
Matches 69; Conservative 9; Mismatches 12; Indels 24; Gaps 6;

QY 2 RFSRSAPPPAYQOGQNLVNEINLGRREYDVLDRKRGDPENMGK-PRKRPQEGLYNE 60
DB 54 KFSRSADAPPYQOGQNLVNEINLGRREYDVLDRKRGDPENMGKPKQRKPNQEGLYNV 113
QY 61 LQDK-----KMAEYSEIGMKGE-----RRRG--KG-----HDGLYQGL 92
DB 114 SRDLTFDGLGKLEGLLE--GSRGAGAGGAGGLQKRWKRGWRAPRPEGLLEGL 166

```

```

RESULT 8
Q7SYB5 PRELIMINARY; PRT; 612 AA.
AC Q7SYB5;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Similar to differentially expressed in FDCP 6.
GN ORFNames=zgc:63721;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=whole body;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Cavaletto T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richard S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=whole body;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; BC054935; AAH54935.1; -.
DR HSSP; O08967; 1FGV.
DR ZFIN; ZDB-GENE-040426-1246; zgc:63721.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF_Hand_1like.
DR InterPro; IPR001849; PH.
DR InterPro; IPR01036; PH-related.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
SQ SEQUENCE 612 AA; 72090 MW; 778BDBC5FBBB689 CRC64;

Query Match 15.3%; Score 90.5; DB 2; Length 612;
Best Local Similarity 28.3%; Pred. No. 4.2;
Matches 28; Conservative 25; Mismatches 43; Indels 3; Gaps 3;

QY 12 YQOGQNLVNEINLGRREYDVLDRKRGDPENMGKPRKRPQEGLYNEQDKKAAEAY 70
DB 309 YVGGKSLIKDKLTKRRDRQREKREKKEBELQRLRLQGERRKMAEELILKKAQNG 368
QY 71 SEIGM-KGERRRGKGDGLYQGLSTATYKDTYDA-LHMQA 107
DB 369 AQAMLEQDEORRRQGHQHLQALQLTQKAEBAASMQA 407

RESULT 9
Q6TNUB

```

```

ID 06TNU8      PRELIMINARY;      PRT;      628 AA.
AC 06TNU8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Differentially expressed in FDCP 6-like protein.
GN Name=DF6; ORFNames=zgc:63721;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Sheng Y., Chen Y., Ruan Z., Jiang C.L., Fan H.Y., Zou L.I.,
RA Kanki J.P., Look A.T., Chen Z.;
RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL: AY391419; AAC91231.1; -.
DR ZFIN: ZDB-GENE-040426-1246; zgc:63721.
DR GO: GO:0005509; F:calcium ion binding; IEA.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR010983; EF_Hand_1like.
DR InterPro: IPR001849; PH.
DR InterPro: IPR011036; PH_related.
DR Pfam: PF00169; PH_1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
SQ SEQUENCE      628 AA; 74478 MW; CF69DCA28BA944F0 CRC64;

Query Match      15.3%; Score 90.5; DB 2; Length 628;
Best Local Similarity 28.3%; Pred. No. 4.3;
Matches 28; Conservative 23; Mismatches 43; Indels 3; Gaps 3;

QY 12 YQGGONQLYNEINIGRREBYDVLKRR-GRDPEMGKPRRKNOEGLYNELQDKMAEAY 70
DB 309 YVSGKSLHMDLTKRRDRQREGRKEAKEQELQRLALQSERRERMALEILKEAQRO 368
DB 369 AQAMLEQDEQRHROQHOLHQLALEIQLEAEARASQQA 407

RESULT 10
ID WCI_NEUCR      STANDARD;      PRT;      1167 AA.
AC 001371;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE White collar 1 protein (WCI1).
GN Name=WC-1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxId=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=74-OR23-1A / FGSC 987;
RA MEDLINE=96203083; PubMed=8612589;
RA Ballario P., Valtorioso P., Magrelli A., Talora C., Cabibbo A.,
RA Macino G.;
RA Neurospora, a central regulator of blue light responses in
RT "White collar-1, a central regulator of blue light responses in
RT Neurospora, is a zinc finger protein.";
RL EMBO J. 15:1650-1657(1996).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Ballario P.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May function as a transcription factor involved in light
CC regulation. Binds and affects blue light regulation of the AL-3
CC gene. WCI1 and WCI2 proteins interact via homologous PAS domains,

```

```

CC bind to promoters of light regulated genes such as FRQ, and
CC activate transcription.
CC -1- SUBUNIT: Heterodimer of WCI1 and WCI2 (potential).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: By blue light.
CC -1- DOMAIN: The glutamine-rich domain might function in activating
CC gene expression.
CC -1- SIMILARITY: Contains 1 GATA-type zinc finger.
CC -1- SIMILARITY: Contains 3 PAS (PER-ARNT-SIM) dimerization domains.
CC -1- SIMILARITY: Contains 2 PAS-associated C-terminal (PAC) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: X94300; CAA63964.2; -.
DR HSSP: P1679; 1GNF.
DR TRASPAC: T02819; -.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR000679; ZnF_GATA.
DR Pfam: PF00320; GATA; 1.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR SMART: SM00086; PAC; 2.
DR SMART: SM00091; PAS; 3.
DR SMART: SM00401; ZnF_GATA; 1.
DR TIGRFAMs: TIGR00229; sensory box; 3.
DR PROSITE: PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE: PS01114; GATA_ZN_FINGER_2; 1.
DR PROSITE: PS50112; PAS; 3.
DR KX Activator; DNA-binding; Nuclear protein; Repeat;
KM Transcription regulation; zinc-finger.
FT DOMAIN 16 61
FT DOMAIN 381 452
FT DOMAIN 469 508
FT DOMAIN 574 644
FT DOMAIN 650 691
FT DOMAIN 693 763
FT ZN_FING 934 959
FT DOMAIN 21 57
FT DOMAIN 329 333
SQ SEQUENCE      1167 AA; 127454 MW; 6489D04DAB50EB38 CRC64;

Query Match      14.6%; Score 86; DB 1; Length 1167;
Best Local Similarity 24.8%; Pred. No. 24;
Matches 33; Conservative 19; Mismatches 49; Indels 32; Gaps 6;

QY 9 PPAYQGGONQLY-----NELNIGRREBYDVLKRRGRDPEMGKPRR-----KX 52
DB 71 PPTTNGNGNSTIHASDVMTSGSDSLDIQLQNDKRRRRSVQPPYGGQRRRLSMFDYAN 130
QY 53 PQEGLYNELOKDMAAYSE-----LCKGERRRGK-----HGQLQGLS---TATK 97
DB 131 PNDG-FSDYQLDMSGNYGDMTGGMGSHSPYAGQNTIAMS DSHGSGYSHMS PNTWGMN 189
QY 98 DTYDALHMQALPP 110
DB 190 MTPYNLMTYHSP 202

RESULT 11
ID Q7RVA7      PRELIMINARY;      PRT;      1262 AA.
AC Q7RVA7;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE White collar 1 protein (WCI1).

```

```

GN Name=NCU02356.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCB1_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh R., Ma L.-J., Smirnov S., Purcell S., Rehm B.,
RA Elkins T., Engels S., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Imanaka C.P., Pedersen D., Nelson M., Washburne M.,
RA Seitzemerkhoff C.P., Kinsey J.A., Brann E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte B., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gierke S.,
RA Kamal M., Kamysseis M., Mauceli E., Bielke C., Rudd S., Fishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cognoli C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.C., Glaes L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander B.S., Nussbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0 (2003).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABX0100358; EAA30541.1; -.
CC HSSP; P16779; IGNF.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR Pfam; PF00785; PAC; 1.
DR TIGRPFAM; TIGR00229; sensory_box; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
DR PROSITE; PS50112; PAS; 3.
DR SSQUENCE 1262 AA; 138115 MW; F39E72B09DE1E5F1 CRC64;
SQ

Query Match 14.6%; Score 86; DB 2; Length 1262;
Best Local Similarity 24.8%; Pred. No. 26;
Matches 33; Conservative 19; Mismatches 49; Indels 32; Gaps 6;

```

```

DE (Hypothetical protein SMU.1408c) (Hypothetical protein
DE SMU.1894c).
GN OrderedLocustName=SMU.1379, SMU.1408c, SMU.1894c, SMU.766;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCB1_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0A159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; Pubmed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang U.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AE014901; AAN58307.1; -.
DR EMBL; AE014918; AAN58486.1; -.
DR EMBL; AE014971; AAN59047.1; -.
DR EMBL; AE014974; AAN59073.1; -.
DR EMBL; AE015015; AAN59507.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004803; F:transposase activity; IEA.
DR GO; GO:0006313; P:DNA transposition; IEA.
DR InterPro; IPR009057; Homodomain_1like.
DR InterPro; IPR002514; Transposase_8.
DR Pfam; PF01527; Transposase_8; 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 172 AA; 20622 MW; 5187DC04B8407F52 CRC64;
SQ

Query Match 14.4%; Score 85; DB 2; Length 172;
Best Local Similarity 36.8%; Pred. No. 3.1;
Matches 14; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

```

```

QY 27 REEYDVLDKRRGRDPENKGRKRRKPPQEGLYNELOKD 64
DB 100 KKNRYTVYKTRGRPPKGRKPKKSEMTLELOKE 137
ID 0754G7 PRELIMINARY; PRT; 744 AA.
AC 0754G7;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE AFR103WP.
GN Name=AFR103W;
OS Ashbya gossypii (yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCB1_TaxID=33169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10895;
RX Pubmed=15001715; DOI=10.1126/science.1095781;
RA Dietrich F.S., Voegel S., Brachat S., Lerch A., Gates K., Steiner S.,
RA Mohr C., Pohlmann R., Luedi P., Choi S., Wing R.A., Flavler A.,
RA Gaffney T.D., Philippsen P.;
RT "The Ashbya gossypii genome as a tool for mapping the ancient
RT Saccharomycetes cerevisiae genome.";
RL Science 304:304-307 (2004).
DR EMBL; AE016901; AAS53474.1; -.
DR AGD; AFR103W; -.
DR InterPro; IPR010983; EF_Hand_1like.
DR InterPro; IPR000261; EFS15_homology.
DR SMART; SM00027; EH; 1.
DR PROSITE; PS50031; EH; 1.
DR SSQUENCE 744 AA; 82630 MW; 7DBE05484729D7D1 CRC64;
SQ

Query Match 13.8%; Score 81.5; DB 2; Length 744;

```


Best Local Similarity 29.3%; Pred. No. 38;
Matches 29; Conservative 17; Mismatches 38; Indels 15; Gaps 5;

QY 4 SRSAPPAVQGGONQYNEINLGRREYVDLDRGRDP-EMGGRPRKPOGGT-NEL 61
DB 527 SSSSEP-----HQSDS-SESEEDYSDNRSSRSRRSGKREKIRQMLRYRNS 578
QY 62 OKDKMAVAYSEIGMKERRRGKHDGLYQGLSTATDITY 100
DB 579 NEDKPMKSHLDVGYTERERKR-----YEGMWVSNNDTY 612

RESULT 14
Q7XK73 PRELIMINARY; PRT; 1123 AA.
AC Q7XK73: 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DE 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
OS OSLJBA0028M15.14 protein.
GN Name=OSJBA0028M15.14;
OC Oryza sativa (Japanese cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
NCBI_Taxid=39947;
[1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Wang Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Liu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.W., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Mu M., Zhang R.O., Guan J.P., Gu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Chen S.T., Ni L., Zhu F.H., Hong G.F.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.,
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL662961; CAB05822.1; -
DR Gramene; Q7XK73; -
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR008916; Retrotrans_gag.
DR InterPro; IPR005162; Retrotrans_gag.
DR InterPro; IPR000477; Retrov_capsid_C.
DR Pfam; PF03732; Retrotrans_gag; 1.
DR Pfam; PF00078; RVT_1; 1.
KM RNA-directed DNA polymerase; Transferase.
SO SEQUENCE 1123 AA; 125576 MW; 6BBF199DFB92A32 CRC64;

Query Match 13.7%; Score 81; DB 2; Length 1123;
Best Local Similarity 26.6%; Pred. No. 69;
Matches 34; Conservative 13; Mismatches 45; Indels 36; Gaps 6;

QY 5 RSAPPAVQGGONQYNEINLGRREYVDL-----KRGDPKMGKPRKRNQ 54
DB 356 RQATPP--PRGTSIDLHLN-GRREARRTRDNRNRHRYVSSRRHNEGGGLSENQH 412
QY 55 EGLYNELQDKMAVAYSEIGMKERRRG-----KGHDGLYQGLSTATDITY-- 100
DB 413 DNRHNRREHDN-----REQRMFGDTRRGCRANDDDGDCRDNGLMQGFVANFGQTYKR 467
QY 101 ----DALH 104
DB 468 HTIEDDLH 475

RESULT 15
Q19579 PRELIMINARY; PRT; 692 AA.
AC Q19579; 019579;

DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Hypothetical protein F18H3.3a.
GN Name=pab-2; Synonyms=F18H3.3a; ORFNames=F18H3.3;
OS Caenorhabditis elegans.
OC Rhabditida; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
NCBI_Taxid=6239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Colas L.;
RT Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z50110; CA80444.1; -
DR F1R; T21095; T21095.
DR HSSP; P11940; 1CIVJ.
DR WormBase; WBGene0003903; pab-2.
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR002004; PABP/HD.
DR InterPro; IPR006515; PABP_1234.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00658; PABP; 1.
DR Pfam; PF00076; RRM_1; 4.
DR SMART; SM00517; PolyA; 1.
DR SMART; SM00360; RRM; 4.
DR TIGRfams; TIGR01628; PABP-1234; 1.
DR PROSITE; PS0102; RRM; 4.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_3.
KM Hypothetical protein.
SO SEQUENCE 692 AA; 75982 MW; D86D7391304AE773 CRC64;

Query Match 13.5%; Score 80; DB 2; Length 692;
Best Local Similarity 21.7%; Pred. No. 48;
Matches 30; Conservative 22; Mismatches 40; Indels 46; Gaps 6;

QY 9 PPAVQGGONQYNEINLGRREYVDLDRGRDP----- 42
DB 486 PPGQRFQNMFMQYPFNQRYV--PQGGRRPWRKTDGRDQYGMAPPAGAPRVGGG 542
QY 43 -EMGGRPRKN-----POEGLYNELQDKMAVAYSEIGMKERRRGKHDGL-YQGLS 93
DB 543 VOMGAPLIRQGPAPRAGAPQKPPYQPPRQDPQHSQPPAQPPGQGGGSGIVIHGE 602
QY 94 TATKDYDALHM--QALP 109
DB 603 TLTS-----HMLAQAP 614

Search completed: March 7, 2005, 07:19:37
Job time : 45.1298 secs

This Page Blank (uspto)